1-2018

Wavelet ANOVA Bisection Method for Identifying Simulation Model Bias

Andrew D. Atkinson

Raymond R. Hill
Air Force Institute of Technology

Joseph J. Pignatiello Jr.
Air Force Institute of Technology

G. Geoffrey Vining
Virginia Tech

Edward D. White
Air Force Institute of Technology

See next page for additional authors

Recommended Citation

Atkinson, A. D., Hill, R. R., Pignatiello Jr., J. J., Vining, G. G., White, E. D., & Chicken, E. (2018). Wavelet ANOVA bisection method for identifying simulation model bias. Simulation Modelling Practice and Theory, 80(January), 66–74. https://doi.org/10.1016/j.simpat.2017.10.002

Authors
Andrew D. Atkinson, Raymond R. Hill, Joseph J. Pignatiello Jr., G. Geoffrey Vining, Edward D. White, and Eric Chicken

Wavelet ANOVA Bisection Method for Identifying Simulation Model Bias

Andrew D. Atkinson[a], Raymond R. Hill[a], Joseph J. Pignatiello, Jr.[a], G. Geoffrey Vining[b], Edward D. White[c], Eric Chicken[d]

[a]*Department of Operational Sciences, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson AFB, OH 45433, USA*
[b]*Department of Statistics, Virginia Tech, 250 Drillfield Drive, Blacksburg, VA 24061, USA*
[c]*Department of Mathematics and Statistics, Air Force Institute of Technology, 2950 Hobson Way, Wright-Patterson AFB, OH 45433, USA*
[d]*Department of Statistics, Florida State University, 117 N. Woodward Ave, Tallahassee, FL 32306, USA*

Abstract

High-resolution computer models can simulate complex systems and processes in order to evaluate a solution quickly and inexpensively. Many simulation models produce dynamic functional output, such as a set of time-series data generated during a process. These computer models require verification and validation (V&V) to assess the correctness of these simulations. In particular, the model validation effort evaluates if the model is an appropriate representation of the real-world system that it is meant to simulate. However, when assessing a model capable of generating functional output, it is useful to learn more than simply whether the model is valid or invalid. Specifically, if the model is deemed invalid, then what aspects of the model are incorrect? Is it possible to identify over what range the model data are a poor representation of the system data? Current V&V methods cannot identify these ranges. This paper proposes a wavelet analysis of variance (WANOVA) bisection method that first assesses model validity and can also identify the interval(s) over which the model is biased. The technique is illustrated using several simulation studies. Ultimately, this new method supports and expands the efficacy of model validation efforts.

Email addresses: `andrew.atkinson@afit.edu` (Andrew D. Atkinson), `raymond.hill@afit.edu` (Raymond R. Hill)

Preprint submitted to Simulation Modelling Practice and Theory *October 3, 2017*

Keywords: Model validation, wavelets, WANOVA, bisection

1. Introduction

Advances in computer hardware technology have allowed the scientific community to build high-resolution computer models capable of simulating complex systems and processes. These computer models can not only evaluate a solution quickly and inexpensively, but also produce dynamic functional output, such as a set of time-series data generated during a process. Since computer simulation technology has quickly advanced, it is critical that the set of verification and validation (V&V) techniques similarly progresses. V&V is an integral part of the simulation development process, one that assesses the accuracy and suitability of the model before relying upon the results.

V&V techniques vary both in quality and applicability to certain models. Often, the quality of the technique may be judged by the amount of subjectivity involved. Basic V&V approaches [1] include subjective, visual comparisons of system data to model data. More advanced methods [2] utilize statistical comparisons of the data that are very complete and more objective. The applicability of a particular V&V technique may depend on the nature of the simulation output. For example, simulation output may include discrete forms and functional forms depending on the system being modeled. Discrete simulation output includes measures such as means and variances, while functional output includes time-series data.

It is clear that while there are a wide variety of V&V techniques available, it is important to select an approach that meets both quality and applicability requirements. This paper focuses on objective, statistical validation techniques used to evaluate models that generate functional output. There are several types of validation methods that meet this criteria [3, 4, 5, 6, 7]. However, once these validation techniques are applied, if the model is assessed as invalid, analysts are still limited in both knowledge and understanding as to the exact nature of the problem leading to the conclusion of an invalid model. The logical, follow-up question to an assessment of invalidity is, "what is wrong with the model?" If the model generates functional output, such as time-series data, it would be very valuable to identify over what range the model data are a poor representation of the system data. Alternatively, over what range is the model data a good representation of the system? Current techniques stop before answering these resulting questions.

This paper presents a sequential validation methodology that helps answer the resulting questions associated with an invalid model based on functional output. This method first assesses the validity of a model using wavelet analysis of variance (WANOVA). If the model is declared invalid, the wavelet-based test statistic is used in conjunction with a traditional bisection univariate search approach to compare the system and model data and identify the interval with the largest discrepancy. This establishes the region in the signal over which the model data are most biased in relation to the system data. The identification of this biased region in the signal then allows developers to correct the appropriate components of the model.

The paper is organized as follows: Section 2 surveys the available literature on model validation and wavelet-based functional data analysis. Section 3 reviews wavelet analysis and WANOVA as a model validation technique. Section 4 presents the WANOVA Bisection method for identifying simulation model bias. Section 5 provides a detailed example of the method applied to a simulation study and the results from a large number of simulations. Finally, Section 6 identifies several distinct invalid model scenarios and assesses the performance of the algorithm under these conditions.

2. Literature Review

The concept of simulation can be traced back to sampling theory demonstrated with the Buffon Needle Experiment in 1777 in what would become the Monte Carlo simulation method [8]. Since then, the advent of computer technology opened new doors in the field of computer simulation. In 1943, Ulam used one of the first electronic general-purpose computers to conduct computer based simulations that would numerically estimate solutions to intractable problems associated with the Manhattan Project and actually coined the phrase Monte Carlo for the statistical sampling approach [8, 9]. With the rise of computer based simulations, some recognized the need to assess the simulation process critically and define a framework of steps to follow to ensure the quality of the resulting simulation. These steps included evaluating the model for both correctness and suitability. In 1979, Sargent [10] presented one of the first in a sequence of papers on simulation validation. Over time, Sargent, Balci [1], and Kleijnen [11] developed some of the foundational work on simulation validation. Today, Balci [1] describes verification as "building the model right," whereas validation evaluates "building the right model."

Over the years, a wide range of validation techniques have emerged. For example, Balci [2] describes informal techniques that rely on human judgment and dynamic techniques that utilize statistical analysis such as hypothesis testing and confidence intervals. However, one needs to recognize that many established statistical techniques are designed for use with models that generate discrete output. Alternative techniques are required to assess models that generate functional output, such as time-series data. Performing analysis on a single parameter, such as the mean, of the functional data is an oversimplification of the system and model results.

Model validation metrics provide a comprehensive technique for evaluating models that generate time-series data. Validation metrics measure the discrepancy between system and model data by calculating the error associated with different signal components, such as correlation, lag, and magnitude. Together, these errors comprise an overall validation metric that describes the level of agreement between two data signals. Oberkampf and Barone [7] discuss the construction of validation metrics and some recommended features. Several authors including Atkinson *et al.* [3], Geers [12], Russell [13], and Sarin *et al.* [14] introduce different versions of validation metrics. However, an important shortcoming with the use of validation metrics is that they still require a subjectively chosen metric value to declare model validity. Accordingly, Sargent [15] expresses concerns with the use of validation metrics and the subjectivity required in their use.

More objective model validation techniques exist within the field of functional data analysis. Functional data analysis is the statistical study of functional data and includes Functional Analysis of Variance (FANOVA). Ramsay and Silverman [16] describe FANOVA as a statistical test on whether a treatment has an effect on the functional response. For time-series data, this basic FANOVA method evaluates a univariate ANOVA for each value of time. Unfortunately, a drawback to this approach is that the dimension of the response can lead to a large number of hypothesis tests and a compounding Type I error rate. Other authors [5, 6, 17] have introduced methods to control this Type I error via multivariate statistics and wavelet thresholding. Wavelets may offer benefits in this regard, as they are known for their data compression capabilities.

Wavelets transform data from the time domain to the time-frequency domain. They offer the benefits of smoothing, dimension reduction, and decorrelation of data [18, 19, 20]. Several authors [17, 21, 22] explore wavelet-based functional data analysis, or WANOVA, an approach whose models operate by transforming the data to the wavelet domain and calculating an appropriate test statistic. This test statistic is applied to general tests of significance with functional data. Section 3 of this paper discusses the dynamics of wavelet analysis in further detail.

Wavelet-based model validation is well-suited for assessing models that generate functional data for the reasons given above. There are wavelet validation methods based on model validation metrics [3, 23] and wavelet coherence [24]. Recently, a WANOVA based validation effort has been proposed [4] which uses a WANOVA test statistic [17] to test for a statistically significant difference between system and model data. This technique offers an objective evaluation of the model that is capable of examining data of large dimension. However, if any of the aforementioned techniques conclude that the model is invalid, there is still little to no information regarding the extent or location of the disparity, nor any insight for correcting the model. A technique that not only assesses model validity but also provides information on the region of model bias would be quite valuable and is presented in this work.

3. Wavelet Analysis and WANOVA

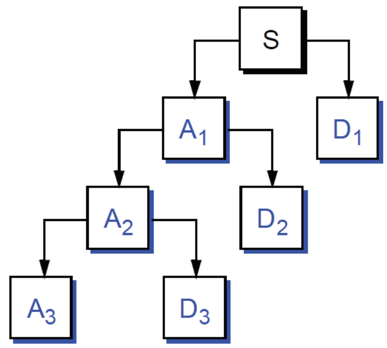
3.1. Wavelets

As introduced above, wavelet analysis transforms data signals from the time domain to the time-frequency domain via a set of wavelet basis functions. Ogden [20] states, "broadly defined, a wavelet is simply a wavy function carefully constructed as to have certain mathematical properties. An entire set of wavelets is constructed from a single 'mother wavelet' function, and this set provides useful 'building block' functions that can be used to describe any in a large class of functions." Wavelets operate in a manner similar to a Fourier transform, but add several advantages, including computational efficiency and the ability to transform non-stationary data. Several textbooks on wavelets [18, 19, 20, 25, 26] are available for further instruction.

Wavelets are typically defined using a mother wavelet (ψ) and father wavelet (ϕ). They may be chosen from a group of established and commonly used wavelets, such as those developed by Daubechies, Meyer, or Shannon. The parent wavelet functions generate an entire family of wavelets through dilations and translations. The dilation index or scale factor is expressed using subscript j and the translation index or shift factor with subscript k. A linear combination of these shifted and scaled versions of the wavelet functions represent a signal as,

$$f(t) = \sum_{k} c_{j_0,k}\phi_{j_0,k} + \sum_{j \geq j_0} \sum_{k} d_{j,k}\psi_{j,k} \ , \tag{1}$$

where $c_{j,k}$ and $d_{j,k}$ represent the wavelet coefficients. These wavelet coefficients are estimated via the Discrete Wavelet Transform (DWT), which calculates the



$$S = A_1 + D_1$$
$$= A_2 + D_2 + D_1$$
$$= A_3 + D_3 + D_2 + D_1$$

Figure 1: Decomposition of signal S into Approximation and Details [27]

inner products of the signal and wavelet functions. In Equation 1, the summation containing the father wavelet is known as the low-frequency "Approximation," while the summation containing the mother wavelet is the high-frequency "Detail."

The wavelet decomposition process separates the high and low frequency content of the signal through an iterative application of the DWT. Each level of signal approximation is successively decomposed into another level of approximation and details until the desired resolution level is attained. Therefore, the level i approximation (A_i) would be decomposed into the level $i+1$ approximation and details. Figure 1 illustrates this process, where A_3 is $\sum_k c_{j_0,k}\phi_{j_0,k}$, D_3 is $\sum_k d_{j_0,k}\psi_{j_0,k}$, D_2 is $\sum_k d_{j_0+1,k}\psi_{j_0+1,k}$, and D_1 is $\sum_k d_{j_0+2,k}\psi_{j_0+2,k}$. Further, the inverse wavelet transform may be applied to perfectly reconstruct the original signal from these approximations and details. The wavelet decomposition process plays a role in the data compression and de-noising capabilities of wavelets, most notably in a process called wavelet thresholding.

Wavelet thresholding, or wavelet shrinkage, is a technique used to compress or de-noise a data signal. It utilizes two key properties of wavelet transforms: orthogonality and sparsity. First, the DWT is an orthogonal linear transform matrix such that the original observed errors are transformed into noisy estimated wavelet coefficients. Second, wavelet sparsity asserts that most of a clean signal's energy is concentrated in a small subset of wavelet coefficients and the remaining coefficients are zero. Donoho and Johnstone [28] introduce thresholding while assuming normal, independent errors. In this case, orthogonality ensures the wavelet-transformed errors retain their normality. They define a universal threshold,

$$\lambda = \hat{\sigma}\sqrt{2log(n)}, \tag{2}$$

where $\hat{\sigma}$ is a consistent estimate of the standard deviation of the noise and n is the sample size. This universal threshold represents a reasonable upper bound on wavelet coefficient noise, so all wavelet coefficients that fall below this threshold are set to zero. The de-noised signal may then be reconstructed with the remaining nonzero wavelet coefficients. Wavelet thresholding is thus an effective technique for de-noising and dimension reduction, and is also an important step in the WANOVA process discussed in the next section.

3.2. WANOVA

WANOVA consists of a statistical hypothesis test performed in the wavelet domain. These wavelet-based ANOVA models offer several benefits, such as smoothing the functional data and reducing dimensionality. Several authors [5, 6, 17, 21, 22, 29, 30] introduce work on WANOVA or related topics. Specifically, this paper advances the methods presented by Girimurugan *et al.* [17] and Atkinson *et al.* [4].

Girimurugan *et al.* develop a WANOVA procedure for detecting differences among functional data aimed at a statistical process control method called profile monitoring. The authors first develop a FANOVA model based on the multivariate Hotelling T^2 statistic. This model considers a functional response Y_{ijk}, for treatment, $i = 1, 2, ..., t$, replicate, $j = 1, 2, ..., r_i$, and response, $k = 1, 2, ..., n$. The noise is assumed multivariate normal, $N(\mathbf{0}, \mathbf{\Sigma})$, with covariance matrix $\mathbf{\Sigma}$ equal to $\sigma^2\boldsymbol{I} \in \mathbb{R}^{n\times n}$ [17].

Girimurugan *et al.* modify the Hotelling-FANOVA statistic by estimating the sum of squares in the wavelet domain and the variation using the Median

Absolute Deviation (MAD), $\hat{\beta}^2$, of the finest scale detail coefficients. This results in the modified Hotelling-FANOVA statistic,

$$\vartheta = \left(\hat{\beta}^2(t-1)\right)^{-1} \sum_{i=1}^{t} \frac{1}{\varsigma_i} \mathbb{W}[\overline{Y}_{i.} - \overline{Y}_{..}]' \mathbb{W}[\overline{Y}_{i.} - \overline{Y}_{..}] \,, \tag{3}$$

where $\varsigma_i = \frac{1}{r_i} + \frac{1}{t}\sum_{j=1}^{t}\frac{1}{r_j}$, $\mathbb{W}$ represents the DWT, subscript "." represents the sum across the applicable parameter, and an overbar represents an average. Then let the set of wavelet coefficients for the treatment i effect be defined by

$$\Theta_i = \hat{\sigma}^{-1}\mathbb{W}\left[\overline{Y}_{i.} - \overline{Y}_{..}\right], \tag{4}$$

and let $\tilde{\Theta}_i = \{\tilde{\theta}_{i1}, \tilde{\theta}_{i2}, ..., \tilde{\theta}_{iT_i}\}$ for the T_i coefficients after thresholding represent the thresholded version of these coefficients. The proposed test statistic,

$$\kappa_\eta = \sum_{i=1}^{t}\sum_{k=1}^{T_i} \tilde{\theta}_{ik}^2, \tag{5}$$

is used to test the null hypothesis that the set of t profiles corresponding to different treatments is statistically equivalent [17].

Atkinson *et al.* [4] adapt this WANOVA methodology to solve model validation problems. The system data signal, $\mathbf{s}$, is compared to the model data signal, $\mathbf{m}$, each with dimension n. If the model is valid, the model data signal should match the system data signal. WANOVA tests the hypotheses that,

$$H_0 : \mathbf{s} = \mathbf{m}$$
$$H_1 : \mathbf{s} \neq \mathbf{m}.$$

The test statistic, κ_η, is nonnegative, and at the α level of significance we reject the null hypothesis if the statistic exceeds a critical value,

$$\kappa_\eta \geq \kappa_\eta(\alpha). \tag{6}$$

Otherwise, we fail to reject the null hypothesis that the model is valid. Under the null hypothesis, the κ_η test statistic is distributed as a χ^2 distribution convolved with a reverse truncated χ^2 distribution with degrees of freedom based on the signal dimension and the number of thresholded wavelet coefficients. In particular, the distribution of κ_η is

$$\kappa_\eta \sim \chi^2_{n_t} \; * \;]\chi^2_{n-n_t}[_{\lambda}\,, \tag{7}$$

where n is the signal dimension, n_t is the number of Approximation wavelet coefficients not considered for thresholding, $*$ represents the convolution operator, $]\,[$ represents a reverse truncated distribution, and λ is the amount of threshold. Girimurugan *et al.* [31] and Atkinson *et al.* [4] describe the distribution of κ_η under the null and alternative hypotheses in greater detail.

4. WANOVA Bisection Method

The WANOVA validation methodology performs a statistical test on functional system and model data to determine whether the data are statistically equivalent. If a statistically significant difference exists, then the model is deemed invalid. This assessment of invalidity invites questions as to the nature of the difference, such as the scope and location of a discrepancy. The following WANOVA Bisection method answers these questions by identifying the interval over which the model data differ the most from the system data.

The WANOVA Bisection method operates using an iterative application of the WANOVA process. Once the model is assessed as invalid, the system and model data signals are bisected. Next, the WANOVA test statistic is calculated for each half of the signal and compared against each other. The signal half with the larger test statistic value is selected and the procedure is repeated on the selected half. This process continues until a desired interval length is achieved.

The resulting interval represents the most biased region of the model data in relation to the corresponding system data.

The steps below summarize the formal WANOVA Bisection method, which mimics a traditional root-finding bisection search method [32].

- Initialization Step
  - Let $[a_1, b_1]$ be the signal interval and let ℓ be the allowable final interval of uncertainty. Let q be the smallest positive integer such that $(\frac{1}{2})^q \leq \frac{\ell}{b_1 - a_1}$. Let $p = 1$ and proceed to the Main Step.

- Main Step
  1. Let $\lambda_p = \frac{a_p + b_p}{2}$. Perform WANOVA over $[a_p, \lambda_p]$ and $[\lambda_p, b_p]$ to calculate $\kappa_{\eta a}$ and $\kappa_{\eta b}$. If $\kappa_{\eta a} > \kappa_{\eta b}$, proceed to Step 2, else proceed to Step 3.
  2. Let $a_{p+1} = a_p$ and $b_{p+1} = \lambda_p$. Proceed to Step 4.
  3. Let $a_{p+1} = \lambda_p$ and $b_{p+1} = b_p$. Proceed to Step 4.
  4. If $p = q$, stop; the model discrepancy lies in the interval $[a_{q+1}, b_{q+1}]$. Otherwise, replace p by $p + 1$ and repeat Step 1.

The analyst may also seek to identify any region(s) in the signal where there is little to no bias. This region may correspond to valid sections of the model data that do not require correction. To identify this portion of the signal, simply modify Step 1 of the WANOVA Bisection method so that, "If $\kappa_{\eta a} < \kappa_{\eta b}$, proceed to Step 2, else proceed to Step 3."

5. Simulation Study

5.1. Example of Method

A detailed example illustrates the WANOVA Bisection method for identifying the region of greatest model discrepancy. Through simulation, we generate a random signal of dimension 1024 for analysis. A series of cosine waves with randomly selected frequency and phase parameters comprise this random signal,

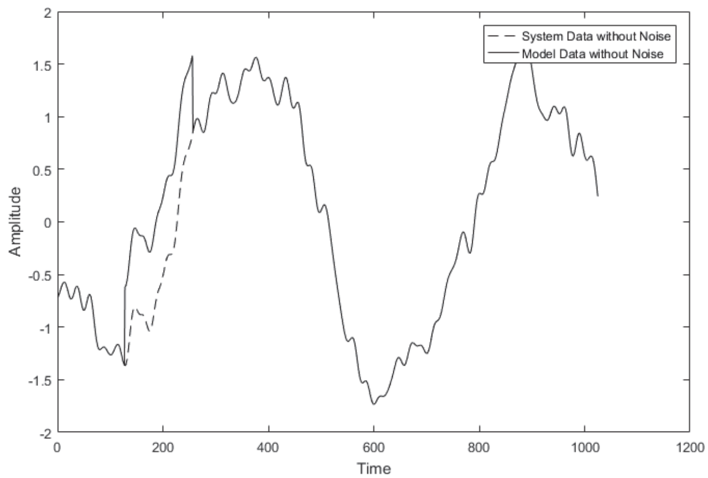


Figure 2: Example Model and System Plot with Biased Region Added

which represents the system data without noise. An additive bias is incorporated into the signal over the interval $[128, 256]$ to represent invalid model data with a specific region of discrepancy. Last, we add normally distributed noise to both the system and model data signals to obtain representative, noisy data. Of note, the noisy signals being analyzed have magnitude ranging from approximately $(-5, 5)$, whereas the model bias is set to $+0.75$. Figure 2 depicts an example of System and biased-Model signals. The signals coincide with the exception of the biased region. Eqach signal then receives its noise component. Figure 3 presents the resulting system (blue) and model data (red) signals. Note that the presence of noise makes it difficult to identify whether the model is invalid, let alone allowing identification of a specific region of model discrepancy due to the bias.

Before applying the WANOVA Bisection method, use WANOVA to assess whether the model is valid. Compare our calculated κ_η test statistic value to a critical value, κ^*_η at the $\alpha = 0.05$ level of significance. We obtain $\kappa_\eta = 207.50 > 144.64 = \kappa^*_\eta$ and therefore reject the null hypothesis and deem the model invalid.

To initialize the WANOVA Bisection method, where $[a_1, b_1] = [0, 1024]$, set

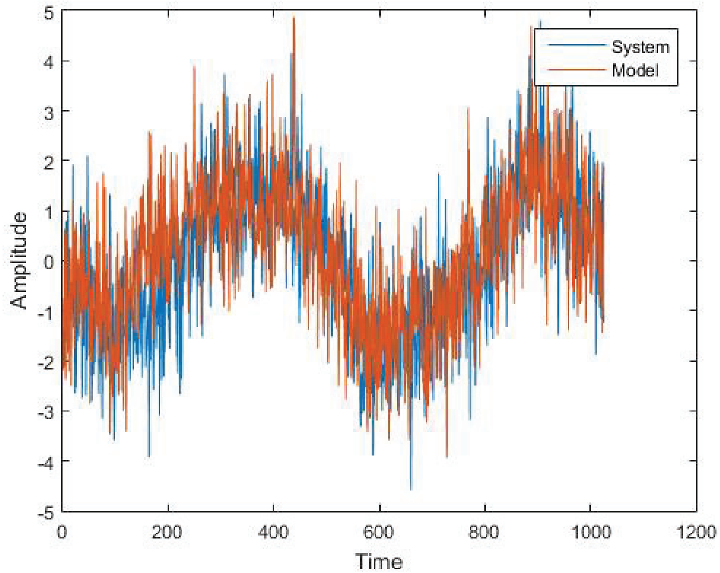


Figure 3: Simulation Study System and Invalid Model Data

$\ell = 128$ as the allowable final interval of uncertainty. In the first iteration of the main step, perform WANOVA over $[0, 512]$ and $[512, 1024]$ to calculate $\kappa_{\eta a}$ and $\kappa_{\eta b}$. In this case, $\kappa_{\eta a} = 129.33 > 88.24 = \kappa_{\eta b}$, so proceed to Step 2 where $[a_2, b_2] = [a_1, \lambda_1] = [0, 512]$. The second iteration performs WANOVA over $[0, 256]$ and $[256, 512]$. Now $\kappa_{\eta a} = 100.41 > 97.46 = \kappa_{\eta b}$, so $[a_3, b_3] = [0, 256]$. In the final iteration, analyze $[0, 128]$ and $[128, 256]$ and calculate $\kappa_{\eta a} = 15.18 < 122.28 = \kappa_{\eta b}$. Therefore, the interval is $[a_4, b_4] = [128, 256]$, and since the algorithm has achieved the allowable final interval of uncertainty, the method stops. The WANOVA Bisection method identifies the interval $[128, 256]$ as the region of largest model discrepancy, which correctly matches the region of inserted bias.

5.2. *An Initial Simulation Study*

To assess the accuracy of the method over a large number of instances, a large number of system and model data signals of length 1024 are simulated. An additive bias equal to 30% of the noisy signal magnitude is incorporated into the model data with a bias duration of 128. Additionally, the system and model data signals include normally distributed noise according to $N(0, 1)$.

The WANOVA Bisection method is applied to 500 instances where the WANOVA null hypothesis is rejected. The allowable final interval of uncertainty is set to 128. Over the 500 instances, the bisection method correctly assesses the interval of discrepancy 491 times, for an overall accuracy rate of 98.2%. The nine incorrect assessments occur when the noise overpowered the bias in the model signal. Nonetheless, the overall accuracy supports the effectiveness of the WANOVA Bisection method at identifying the region of model discrepancy.

The results obtained are positive, but not fully conclusive. Overall, the scenarios considered are relatively benign. The bias inserted into the model signal falls completely within the allowable final interval of uncertainty, and both the bias duration and allowable interval of uncertainty have length 128. Furthermore, the bias is only inserted in one location in the model signal, which allows the bisection method to focus on that particular region of discrepancy. Section 6 considers more complex scenarios that may be more representative of real-world system and model data. These more challenging biased model conditions serve to demonstrate the effectiveness of the WANOVA Bisection method in a variety of invalid model scenarios.

6. Invalid Model Scenarios

6.1. Incorrect Specification of Interval

The analyses from the previous section presumed knowledge of the model bias duration. In real-world applications the analyst does not have prior knowledge of the duration of the model bias. Thus, it is important to assess the effectiveness of the validation technique when the allowable final interval of uncertainty is incorrectly specified. Towards this purpose, two scenarios are evaluated: the acceptable interval is larger than the actual region of bias, and the acceptable interval is smaller than the actual region of bias.

In the first scenario, the duration of the inserted model bias is 64, but the allowable final interval of uncertainty remains at 128. The study assesses whether the WANOVA Bisection method identifies an interval that contains the smaller region of model bias. All other study parameters remain the same from the previous simulations. Over 500 instances, the bisection method correctly identifies the interval that contains the bias region 94.9% of the time. The slightly lower accuracy rate is attributable to the smaller bias region having a smaller effect compared to the signal noise.

In the second scenario, the duration of the inserted model bias is 192 data points long with an allowable interval of uncertainty of 128. Since the bias now spans two search regions, the search region that is completely biased is the *majority region of bias*, while the search region that contains the remaining 64 biased data points is the *minority region of bias*.

The study evaluates whether the WANOVA Bisection method identifies the interval with the highest proportion of bias, i.e. the majority region of bias. All other study parameters remain the same. Over 500 instances, the bisection method correctly identifies the majority region of bias 77.8% of the time. Additionally, the bisection method identifies the minority region of bias 22.0% of the time. The algorithm was incorrect for 0.2% of iterations. Ideally, the method would identify the majority region of bias more consistently. This is due to the algorithm detecting the bias in the minority region. Nevertheless, overall the method was accurate in identifying one of the bias regions 99.8% of the time. Ultimately, these two scenarios show that despite the incorrect specification of the interval with respect to the true bias duration, the WANOVA Bisection method is still extremely effective at identifying the intervals in the model data that are biased.

6.2. Multiple Bias Regions

The previous analyses assessed whether the bisection method could identify the single region of model discrepancy. In practice, the region of bias may not be limited to just one interval. If there are two or more regions of bias in the model data, it is useful to understand which region the algorithm will identify first and what steps should be taken to find the other areas of invalidity.

The next study analyzes model data with two separate regions of varied model bias, a strong bias region where the bias is equal to 30% of the noisy signal magnitude and a weak bias region where the bias is equal to 15% of the noisy signal magnitude. All other parameters are unchanged. Over 500 instances, the algorithm identifies the strong bias region 86.5% of the time, the weak bias region 6.0% of the time, and was incorrect 7.4% of the time. The results indicate that the algorithm is more likely to identify the more biased region first. Further, a greater bias discrepancy between the two regions results in a higher likelihood that the bisection method will identify the strong bias region. Under similar levels of bias in the two regions, the technique finds each region about half of the time, as appropriate and expected.

The last study retains the two separate bias regions, but varies the bias length in each region. In particular, the long bias region has a duration of 100 data points, while the short bias region has a duration of 50. While the length of the bias is different, the magnitude of bias is equivalent. Over 500 instances, the WANOVA Bisection method identifies the long bias region on 79.8% of iterations, the short bias region on 12.2% of iterations, and an incorrect region on 8.0% of iterations. Overall, the technique generally identifies the region of highest discrepancy first, whether it is due to a larger magnitude of bias or a larger duration of bias. These results align with the preferred order of identification. Table 1 summarizes the results from all five of the computational analyses.

In the majority of analyses, the WANOVA Bisection method correctly identifies the more biased interval first. However, one must also account for any other regions of bias. These other problem areas need to be identified. There are two ways to initially account for this situation.

The first option is to extract the biased interval from the system and model data. Then, re-assess the remaining data signals using the suite of WANOVA model validation techniques we have described. A second option is to inform

Table 1: WANOVA Bisection Summarized Results

Scenario Considered	**Correct Identification of Biased Region**			**Incorrect Identification**
	Primary	*Secondary*	*Total*	
Large Run Study	98.2%	-	98.2%	1.8%
Incorrect Interval (oversized)	94.9%	-	94.9%	5.1%
Incorrect Interval (undersized)	77.8%	22.0%	99.8%	0.2%
Multiple Bias Regions (magnitude)	86.5%	6.0%	92.5%	7.5%
Multiple Bias Regions (duration)	79.8%	12.2%	92.0%	8.0%

the model developers of the originally identified interval of bias allowing them to correct the necessary components of the simulation model. Then re-assess the improved version of the model using the WANOVA model validation technique. Ideally, the improved version will not show the original interval as biased. The analysts and developers can iterate this test-and-fix process using our WANOVA validation methods.

7. Conclusion and Recommendations

This paper presents a new model validation methodology that identifies the interval(s) in model data that are most biased in relation to associated system data. This procedure first evaluates the simulation model by executing a WANOVA validation assessment to determine if the system and model functional data are statistically equivalent. If a statistical difference exists, then the WANOVA Bisection method identifies the region of greatest model discrepancy. It may also be modified to identify the region of least discrepancy. The paper illustrates the approach via several simulation studies which demonstrate that the WANOVA Bisection method is an effective technique for identifying the most biased interval in the model data.

Although the WANOVA Bisection method is very accurate, there are several considerations and limitations. First, the bias in the model data must be significant enough to be detected among the signal noise in the data. Second, the wavelet thresholding and WANOVA technique rely upon the assumption of normally distributed noise. Third, the method identifies the region of model discrepancy but does not provide information on the scope or nature of the bias—positive or negative. Last, as discussed in Section 6, the method performs best when the bias is isolated to a single location in the model data and the final interval of uncertainty is correctly specified.

In spite of these considerations, the WANOVA Bisection method is quite robust and effective under the conditions considered in our initial studies at identifying regions of model discrepancy. The studies in this paper show that a bias as low as 30% of the noisy signal magnitude is sufficient for the algorithm to be correct on over 98% of problem instances. In addition, the method is quite robust to more challenging out-of-control conditions, evidenced by accuracy rates above 92% in a variety of invalid model scenarios.

There are many areas of further investigation. Future work will consider nonparametric solutions to the problem and also develop methods that not only identify the location of model discrepancy, but also provide information on the scope and nature of the bias. Another question is how well the method scales when there are many regions of discrepancy. With many potential discrepancies, does the bisection method remain effective or are other search methods more applicable. Finally, we limited our noise and bias components to a narrow range when in fact there are potentially many types of noise and bias patterns that can be examined with respect to improving the robustness of the methodology.

Based on the studies in this paper, the WANOVA Bisection method is a very effective technique for assessing regions of model discrepancy during the validation process. This validation procedure can objectively evaluate functional system and model data through the WANOVA validation process. If the model is deemed invalid, the methodology helps answer some of the resulting questions that typically arise. In particular, the bisection method calculates the WANOVA test statistic value over different intervals of the data and performs a series of comparisons to identify the region of largest model discrepancy in relation to the system data. This process provides solutions to the questions that current validation techniques fall short of answering. The identification of the biased interval(s) in the model data assists model developers to determine what aspects of the simulation model must be corrected. Thus, the WANOVA Bisection method is a valuable technique for identifying model bias and represents a critical step in the model validation process for functional data output.

This research was supported by the Office of the Secretary of Defense, Director of Operational Test and Evaluation (OSD DOT&E) and the Test Resource Management Center (TRMC) within the Science of Test research consortium.

Disclaimer. The views expressed in this article are those of the authors and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the US Government.

References

[1] O. Balci, Verification, validation, and certification of modeling and simulation applications, in: S. Chick, P. Sanchez, D. Ferrin, D. Morrice (Eds.), Proceedings of the 2003 Winter Simulation Conference, Piscataway, NJ: Institute of Electrical and Electronics Engineers, 2003, pp. 150–158.

[2] O. Balci, Verification, validation, and testing, in: Encyclopedia of Operations Research and Management Science, 3rd Edition, New York: Springer, 2013, pp. 1618–1627.

[3] A. Atkinson, R. Hill, J. Pignatiello, G. Vining, E. White, E. Chicken, Dynamic model validation metric based on wavelet thresholded signals, Journal of Verification, Validation and Uncertainty Quantification 2 (2) (2017) 021002.

[4] A. Atkinson, R. Hill, J. Pignatiello, G. Vining, E. White, E. Chicken, Wavelet anova approach to model validation, Simulation Modelling Practice and Theory (To Appear).

[5] J. Fan, Test of significance based on wavelet thresholding and Neyman's truncation, Journal of the American Statistical Association 91 (434) (1996) 674–688.

[6] J. Fan, S.-K. Lin, Test of significance when data are curves, Journal of the American Statistical Association 93 (443) (1998) 1007–1021.

[7] W. L. Oberkampf, M. F. Barone, Measures of agreement between computation and experiment: validation metrics, Journal of Computational Physics 217 (1) (2006) 5–36.

[8] R. E. Nance, R. G. Sargent, Perspectives on the evolution of simulation, Operations Research 50 (1) (2002) 161–172.

[9] D. Goldsman, R. E. Nance, J. R. Wilson, A brief history of simulation, in: M. Rossetti, R. Hill, B. Johansson, A. Dunkin, R. Ingalls (Eds.), Proceedings of the 2009 Winter Simulation Conference, Austin, TX: Institute of Electrical and Electronics Engineers Computer Society, 2009, pp. 310–313.

[10] R. G. Sargent, Validation of simulation models, in: Proceedings of the 1979 Winter Simulation Conference, San Diego, CA: Institute of Electrical and Electronics Engineers Computer Society, 1979, pp. 497–503.

[11] J. P. Kleijnen, Verification and validation of simulation models, European Journal of Operational Research 82 (1) (1995) 145–162.

[12] T. L. Geers, An objective error measure for the comparison of calculated and measured transient response histories, The Shock and Vibration Bulletin 54 (1984) 99–108.

[13] D. M. Russell, Error measures for comparing transient data: part I: development of a comprehensive error measure, in: Proceedings of the 68th Shock and Vibration Symposium, Hunt Valley, MD: Shock and Vibration Exchange, 1997, pp. 175–184.

[14] H. Sarin, M. Kokkolaras, G. Hulbert, P. Papalambros, S. Barbat, R.-J. Yang, Comparing time histories for validation of simulation models: error measures and metrics, Journal of Dynamic Systems, Measurement, and Control 132 (6) (2010) 061401–1–061401–10.

[15] R. G. Sargent, Verification and validation of simulation models, in: S. Chick, P. Sanchez, D. Ferrin, D. Morrice (Eds.), Proceedings of the 2003 Winter Simulation Conference, Piscataway, NJ: Institute of Electrical and Electronics Engineers, 2003, pp. 37–48.

[16] J. O. Ramsay, B. W. Silverman, Functional Data Analysis, 2nd Edition, New York: Springer Science Business Media, Inc., 2005.

[17] S. Girimurugan, E. Chicken, J. J. Pignatiello Jr, M. S. Zeisset, Wavelet anova for detection of local and global profile changes, in: A. Krishnamurthy, W. Chan (Eds.), Proceedings of the 2013 Industrial and Systems Engineering Research Conference, San Juan, PR: Institute of Industrial Engineers, 2013, pp. 3235–3244.

[18] C. S. Burrus, R. A. Gopinath, H. Guo, Introduction to Wavelets and Wavelet Transforms, 1st Edition, Upper Saddle River, New Jersey: Prentice Hall, Inc., 1998.

[19] C. K. Chui, An Introduction to Wavelets, 1st Edition, Boston, MA: Academic Press, 1992.

[20] T. Ogden, Essential Wavelets for Statistical Applications and Data Analysis, 1st Edition, Boston: Birkhauser, 1997.

[21] J. L. McKay, T. D. Welch, B. Vidakovic, L. H. Ting, Statistically significant contrasts between EMG waveforms revealed using wavelet-based functional anova, Journal of Neurophysiology 109 (2) (2013) 591–602.

[22] B. Vidakovic, Wavelet-based functional data analysis: theory, applications and ramifications, in: T. Kobayashi (Ed.), Proceedings of the 3rd Pacific Symposium on Flow Visualization and Image Processing, Maui, HI, 2001.

[23] Z. Cheng, J. Pellettiere, N. Wright, Wavelet-based test-simulation correlation analysis for the validation of biodynamical modeling, in: Proceedings of the 24th Conference and Exposition on Structural Dynamics, St. Louis, MO, 2006, pp. 2124–2132.

[24] X. Jiang, S. Mahadevan, Wavelet spectrum analysis approach to model validation of dynamic systems, Mechanical Systems and Signal Processing 25 (2) (2011) 575–590.

[25] A.-H. Najmi, Wavelets A Concise Guide, 1st Edition, Baltimore, MD: Johns Hopkins University Press, 2012.

[26] J. S. Walker, A primer on wavelets and their scientific applications, 2nd Edition, Boca Raton, FL: CRC press, 2008.

[27] M. Misiti, Y. Misiti, G. Oppenheim, J.-M. Poggi, Wavelet Toolbox Getting Started Guide, 1st Edition, Natick, MA: The Mathworks, Inc, 1997.

[28] D. L. Donoho, J. M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (3) (1994) 425–455.

[29] F. Abramovich, A. Antoniadis, T. Sapatinas, B. Vidakovic, Optimal testing in a fixed-effects functional analysis of variance model, International Journal of Wavelets, Multiresolution and Information Processing 2 (4) (2004) 323–349.

[30] J. Raz, B. I. Turetsky, Wavelet anova and fMRI, in: M. Unser, A. Aldroubi, A. Laine (Eds.), SPIE's International Symposium on Optical Science, Engineering, and Instrumentation, International Society for Optics and Photonics, 1999, pp. 561–570.

[31] S. B. Girimurugan, Nonlinear multivariate tests for high-dimensional data using wavelets with applications in genomics and engineering, Ph.D. thesis, Florida State University (2014).

[32] M. S. Bazaraa, H. D. Sherali, C. M. Shetty, Nonlinear Programming Theory and Algorithms, 3rd Edition, Hoboken, NJ: John Wiley & Sons, Inc., 2006.